Can Your Toothpaste Shopping Predict Mutual Funds Purchasing? — Transferring Knowledge from Consumer Goods to Financial Products via Machine Learning

Shenghao Wang
Tippie College of Business, University of Iowa, Iowa City, IA 52246, shenghao-wang@uiowa.edu

Tong Wang
Tippie College of Business, University of Iowa, Iowa City, IA 52246, tong-wang@uiowa.edu

Cheng He
Wisconsin School of Business, University of Wisconsin - Madison, Madison, WI 53706, che78@wisc.edu

Yu Jeffrey Hu
Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308, jeffrey.hu@scheller.gatech.edu

In recent years, e-commerce platforms have expanded into an important new product domain of financial products that includes payment, credit, investment, and insurance products. However, due to the scarcity of data in this new product domain, online platforms face challenges in predicting users' purchase behavior. In this paper, we study whether we can "transfer" knowledge learned from the existing consumer goods domain to benefit the prediction in the domain of the financial products. With data provided by one of the largest online shopping platforms in China, we develop machine learning solutions to enable knowledge transfer. We show that users' prior browsing and shopping history in consumer goods can significantly improve the prediction accuracy of users' purchases of mutual funds for both the existing-user and the new-user scenarios. In addition, we study the heterogeneous prediction performance lifts on users with different socioeconomic statuses and investment risk preferences. Results show that information from the consumer goods domain has a higher prediction performance lift on users in the high socioeconomic group. Finally, we compare the effect of different sources of information on predicting users' purchases of mutual funds. We find that users' browsing and shopping history for consumer goods are more predictive than their profile features. Our findings and methods will be valuable to both the financial industry and online platforms that seek to expand their product domains.

Key words: cross-domain recommendation, cross-domain consumer behavior, e-commerce, transfer learning

1. Introduction

Over the past two decades, e-commerce has steadily evolved from selling books, CDs, and electronics to trading any kind of products or services that a user could want. For instance, today, Amazon sells over 75 million unique products over 27 categories (ScrapeHero 2021, Connolly 2021). Leading global e-commerce platforms like Amazon, Alibaba, and JD.com are constantly looking for opportunities to expand into new product domains, and the pandemic has sped up this trend of domain expansion (Pere 2020). Recently, one of the most important product domains that e-commerce platforms expanded into is financial products that are traditionally purchased at offline investment agencies (Xia and Hou 2016, Wang and Yang 2019). For example, Amazon has made several fintech investments, pursuing various initiatives ranging from payment and lending to insurances (CBInsights 2022). JD.com, one of the largest e-commerce platforms in China, established JD Digits in 2013 to sell financial products such as mutual funds, insurances, etc. Similarly, Ant Group, an affiliate company of Alibaba, started to sell a broad spectrum of financial products in 2014 (Kharpal 2021).

However, like any other new product domain, the new financial product domain suffers from the issue of cold start due to the lack of data on users' clicks and purchases. Such data limitations prevent platforms from accurately targeting customers via product recommendations and customized promotion. On the other hand, these e-commerce platforms possess rich data of users' online activities in existing domains such as the consumer goods, the earliest product domain that became available online. Because e-commerce platforms have collected rich data and have a good understanding of consumer behavior in this domain, it is natural to wonder whether information from the consumer goods domain can be borrowed and *transferred* to predict users' purchase decisions in the financial products domain, even though these two product domains are entirely different.

Unfortunately, existing literature provides little guidance on whether such a transfer can be successful. First, while previous studies in the consumer behavior literature have shown that information from users' shopping behavior in one category, such as price sensitivity, can be used to predict their behavior in other categories, these studies are limited to categories in the same domain of consumer goods (Blattberg et al. 1978, Ainslie and Rossi 1998, Wedel and Zhang 2004). Whether a cross-domain transfer between consumer goods and financial products is feasible remains unknown. There exists no prior study establishing

such cross-domain connections, primarily due to the fact that the expansion into financial products has been fairly recent. Second, the machine learning literature has studied cross-domain recommender systems which use deep learning, collaborative filtering or matrix factorization techniques to recommend products to users in a new domain while utilizing information from existing domains. However, existing methods are built for domains with content-level or item-level relevance, such as movies and books (e.g., Fu et al. (2019)), videos and news (e.g., Kanagawa et al. (2019)), videos and games (e.g., Kang et al. (2019)), etc. There have not been any attempts on a "transfer" between two seemingly unrelated domains, i.e., consumer goods and financial products, where there is no content-level or item-level relevance. So we ask, is a *transfer* between these two domains viable?

In addition, the problem becomes even more challenging when target users are new to the platform and have no prior activities in existing domains. Current cross-domain recommender systems for the new-user scenario are not feasible in our context since they need some additional data, such as users' opinions (Shi et al. 2017), social tags (Zhou et al. 2021), or detailed viewed item features (Rajendran and Sundarraj 2021), etc., to characterize a user. These existing methods can be considered as different ways to represent users' product preferences or traits. Unfortunately, in settings like ours, there is no auxiliary data for characterizing a new user. Therefore, it is unclear whether any transfer is possible, and more interestingly, what should be transferred, if it is possible, between the two domains?

To answer the above research questions, we collaborate with one of the largest e-commerce platforms in China, which has rich data on consumer goods and has started selling mutual funds in recent years. We randomly select 10,000 users who have both purchased consumer goods and invested in mutual funds. Our datasets contain clickstream data and transaction data for products in both domains. With this dataset, we investigate whether the "knowledge transfer" from the source domain, consumer goods, to the target domain, financial products, is feasible. More specifically, we examine whether users' information in consumer goods can improve the accuracy of predicting their purchase intention of financial products and what methods should be used. We study this research problem in two scenarios: existing-user scenario and new-user scenario.

We start with the *existing-user scenario*, where the target users in the financial product domain have prior activities in the consumer goods domain. To utilize users' information from users' prior browsing history with consumer goods, we solve the problem of

how to *extract knowledge* from these data. We test two solutions for feature extraction: theory-based manual feature engineering and machine learning assisted user representation learning. Both methods show that the information extracted from users' activities in consumer goods category can significantly improve the performance of predictions for mutual funds purchases by 8.97% and 12.41% in R^2, respectively.

We then investigate the *new-user scenario*, where the target users in the financial product domain do not have any prior activities associated with consumer goods. As current cross-domain recommender systems are not feasible in this context, we propose a new way of knowledge transfer inspired by theories on consumer behavior and psychology. Specifically, we transfer the conversion behavior, i.e. the *mapping* from browsing products to purchasing a product, from the consumer goods domain to the domain of the financial product. Such mapping is also called *conversion funnel* (Jansen and Schuster 2011) in the marketing literature. This way of transfer is motivated by prior studies showing that conversion funnel exhibits similarity across product categories within the consumer goods domain (De Haan et al. 2016). In this study, we further explore whether conversion funnels across two seemingly unrelated domains are similar. Therefore, we first build a model to learn the mapping from users' browsing activities to their purchases in the consumer goods domain. Then we adapt such mapping to mutual funds. This way of transfer is distinct from existing methods that transfer users' product preferences or traits, which are not feasible in our setting. Results show that transferring conversion funnel from consumer goods domain can significantly improve the prediction performance in the mutual funds domain by 14.5% in R^2, which suggests that users' decision-making processes for purchasing in these two seemingly unrelated domains bear some similarities.

In addition, we investigate the heterogeneous model performance lift on users with different characteristics. We find the model performance has a higher lift on users with high socioeconomic status and high risk-seeking preference. Lastly, we compare the model performance lift of using information extracted from users' activities in consumer goods with using their profile features. Results show that information in consumer goods could improve the model prediction performance by 8.97% more than using user profile information.

In summary, our work makes the following contributions. First, this study contributes to the literature on similarities between users' shopping behavior in different product domains. Previous literature has mainly studied consumer behavior across product categories within

the same domain. We generalize the connections to seemingly unrelated domains and show that there still exhibit similarities. Such similarities can be exploited by machine learning to better predict users' purchase intention in a new product domain. To the best of our knowledge, we are the first to study the knowledge transfer from the consumer goods domain to the financial products domain. Therefore, our findings also contribute to the literature of investors' behavior by showing that investors' choices can be traced from their purchases and search history in consumer goods.

Second, our proposed methods for knowledge transfer contribute to the literature on cross-domain predictions. Our proposed solutions inherit the genes from several state-of-the-art machine learning techniques, including deep learning, embedding, representation learning, and transfer learning. To tackle the challenges in the new-user scenario, we propose a new mechanism of transferring *conversion funnel*, i.e., the *mapping* from the browsing behavior to final purchases, which differs from transferring product preferences or user traits that many existing cross-domain recommender systems do. Thus, our proposed methods enable knowledge transfer in challenging scenarios with no content relevance, no item relevance, and even no user overlap. In addition, our solutions can be applied to a more general product domain expansion beyond financial products, since our methods do not require detailed item features.

2. Related Work

In this section, we first review state-of-the-art machine learning methods for cross-domain recommendations and discuss their differences to our solutions. Then, we discuss two sub-domains in machine learning that motivated most cross-domain recommendation systems as well as our methods: representation learning and transfer learning.

2.1. Cross-Domain Recommender Systems

Recommender systems are machine learning algorithms that infer users' preference and suggest related items to users (e.g., movies to watch, text to read, and products to buy). In recent years, as e-commerce industry is rapidly expanding the product catalog, cross-domain recommender systems have received increasing interest. Cross-domain recommender systems utilize knowledge learned from a richer (source) domain to assist recommendations in the sparse (target) domain. Existing literatures regarding cross-domain recommender system can be generally classified into three groups (Zhu et al. 2021): 1) content-based transfer, 2) rating pattern-based transfer and 3) embedding-based transfer.

Content-based transfer identifies similar users or items by creating links based on the common contents, such as item details (Winoto and Tang 2008), user-generated reviews (Tan et al. 2014), and social tags (Fernández-Tobías and Cantador 2014), then transfer knowledge between similar users or items across domains. The content-based relations also exist in user attributes (Berkovsky et al. 2007), semantic properties (Kumar et al. 2014, Zhang et al. 2019), thumps-up (Shapira et al. 2013), text information (Tang et al. 2012, Tan et al. 2014), and browsing or watching history (Elkahky et al. 2015, Kanagawa et al. 2019). In other words, content-based transfer need to build content-based relations in source and target domains, in which concrete contexts are required. The second type, ratting patter-based transfer approaches are in fact a special case of content-based transfer. They first learn an independent rating pattern of users from the source domain and then transfer the rating pattern to the target domain to improve the corresponding recommendation accuracy (Zhu et al. 2021). The rating patterns are discovered using rating data of users (Yuan et al. 2019), which is usually very rich in source domain, but relatively sparse in target domain. Finally, embedding-based transfer approaches utilize item and user information to learn item or user embeddings by training different collaborative filtering-based models, including singular value decomposition (Deerwester et al. 1990), maximum-margin matrix factorization (Srebro et al. 2004), probabilistic matrix factorization (Mnih and Salakhutdinov 2008), neural collaborative filtering (He et al. 2017), and graphic models (Zhao et al. 2020). The learned user or item embeddings will be transferred through common user or items across domains. Both embedding-based and rating patter-based approaches employ machine learning techniques as the model framework, including multi-task learning (Singh and Gordon 2008), transfer learning (Zhang et al. 2016), clustering (Farseev et al. 2017), and neural network (Zhu et al. 2020). In addition to those machine learning techniques, other embedding-based transfer approaches also employ some deep learning methods, such as reinforcement learning (Liu et al. 2018).

A particular critical challenge for cross-domain recommendation is the new user problem, where a new user to the platform does not have prior online activities, such as review, rating, or browsing history. Existing methods dealing with new user cold start issue can be typically summarized into three groups (Son 2016) - 1) Make use of additional data from auxiliary domain, 2) Choose the most prominent groups of analogous users by clustering, 3) Enhance the prediction using hybrid methods. The idea of first type methods is to use

some external data, such as users' opinion (Shi et al. 2017), review (Margaris et al. 2020), ratings (Yu et al. 2019), or social tags (Zhou et al. 2021), to improve the cross-domain recommendation for the new user. But auxiliary information is not always available. The second method utilizes the clustering algorithms (Hafshejani et al. 2018, Liu et al. 2014) to determine the analogous users. However, the optimal number of groups and the splitting criteria is difficult to select. The third approach uses hybrid information, such as browsing history and item features (Elkahky et al. 2015), text and metadata (de Souza Pereira Moreira et al. 2018), or browsing records with topic information (Rajendran and Sundarraj 2021), to calculate user similarity or the prediction of user preference. This method requires specific item information, which may not be available in other domains. In summary, existing approaches regarding the new user cold start problem need rich information in the source domains or auxiliary data to characterize traits or product preferences of users.

2.2. Representation Learning

Representation learning is an important machine learning technique that attempts to obtain latent information from raw input data (Bengio et al. 2013). The goal of representation learning is to transform data from its original form that is too difficult to represent (such as graphs, images, etc.) to a much simpler and more efficient representation while preserving the original information and remaining predictive for the task. Compared to feature engineering that relies on human ingenuity and experience to construct features from the raw data, representation is completely data-driven and optimized, thus often enjoying better performance than theory-based representations (Goodfellow et al. 2016). Representation learning has been accompanied and nourished in several fields, including speech recognition, machine translation (Hinton et al. 2012, Mikolov 2012), object recognition (Hinton et al. 2006, Bengio et al. 2006) and image representation (Krizhevsky et al. 2012).

Representation learning has also been applied in recommendation systems. For example, user embedding learns a representation of users to captures their long-term interest (Sun et al. 2021). For example, the search ranking team in Airbnb (Grbovic and Cheng 2018) modified the Skip-Gram model to conduct the user type embedding by leveraging items and metadata. ClickGraph (Jenkins 2019), which is a recurrent neural network that encodes the graph structure of user click trajectories in the learned representations of web pages, is another embedding approach that simultaneously predicts future user actions given a user's

clickstream history. Besides, the sequence of user actions, such as items that were clicked or purchased (Nedelec et al. 2017), queries and advertisements that were clicked (Grbovic et al. 2015), can be treated as context as well. Singh et al. (2019) introduced similar embedding framework that can process different types of data by learning representations from the catalog text data, user's clickstream data, and product images. Item embedding is also a relatively new deep learning technique that produces embedding vectors for items in latent space when user information is not available. This method is capable of computing item similarity and inferring item relationships. For example, item2vec (Oren and Noam 2016) has been proposed and demonstrated in both quantitative and qualitative evaluations for a recommendation based on item similarity.

2.3. Transfer Learning

Transfer learning is a machine learning technique that adapts knowledge and experience obtained from a source domain to a different but related target domain, thus it has been widely studied. In other words, transfer learning also refers to the situation where what has been learned in one setting is exploited to improve generalization in another setting (Goodfellow et al. 2016). For example, knowledge gained while learning to recognize cars could be applied for recognizing trucks. Transfer learning is motivated by the fact that people can intelligently apply knowledge learned previously to solve new problems that are similar or related to old problems faster or with better solutions (Pan and Yang 2010).

Transfer learning provides a reasonable start point for training the model by leveraging previous learnings, such as model parameters, to avoid starting from scratch. The underlying model will be trained on specific data and then applied to another task or data as a pre-trained model. One of the common applications of transfer learning is image classification (Krizhevsky et al. 2012, He et al. 2016). Many image classification tasks re-use the first a few layers from other image classification models trained on a large benchmark dataset to solve another problem similar to the previous one (Oquab et al. 2014, Zhu et al. 2011). The idea is that lower layers in a trained neural network produce general features that are problem independent, while higher layers produce specific features that are problem-dependent (Yosinski et al. 2014). Depending on the problem formulation, one can transfer different neural network layers into another model by freezing the hidden layers. Examples of transferable layers include convolution layers, fully-connected layers, or embedding layers. Therefore, the low-level or mid-level feature representations to classify

one type of object are also helpful for predicting another type of object. The assumption for transfer learning is that models for related tasks are supposed to share some parameters or prior distributions of hyperparameters. The comprehensive review of transfer learning is available at (Pan and Yang 2010). In this paper, we inherit the idea from the above and transfer hidden layers from the pre-trained model to the new predictive model.

2.4. Distinctions of Our Work

Compared to the cross-domain recommendation systems discussed previously, our work has three major distinctions.

First, prior research mainly transfers across similar or relevant domains, while our paper studies the transfer across two entirely different domains. For example, Fu et al. (2019), Yu et al. (2019), Taneja and Arora (2018), Li and Tuzhilin (2020) transfer between movies to books since the two domains can be related by topics, keywords, or writers. Kang et al. (2019) transfers from videos to games. Cao et al. (2010) considers items in different sub-categories as different domains (e.g., textbooks and novels). Some works transfer within exactly the same domain but on different systems such as movie ratings on Douban vs IMDB (Man et al. 2017), MovieLens vs EachMovie (Gao et al. 2013), or different video streaming sites (Gao et al. 2019). In all the previously mentioned works, the source and target domains are very similar or at least partially related by their contents. This is because many existing method need to have content-relevance (e.g., related features such as topics, keywords, etc) or item-relevance (e.g., common items) between the source domain and target domain to work. In our context, however, we aim to transfers knowledge across completely different domains without any content-level relevance. To the best of our knowledge, it is the first attempt to explore the possibility of knowledge transfer from consumer goods to mutual funds and provide solutions for the transfer.

Second, the existing approaches often need detailed item features or content features in order to characterize users. For example, content-based transfer relies on metadata that are often limited to scenarios where the source and target domains have detailed content information, such as movies (Zhu et al. 2020), music and book (Fu et al. 2019, Zhao et al. 2020), or research publications and authors (Tang et al. 2012), etc. Embedding-based and rating patter-based approaches require either user-level relevance or item-level relevance (i.e., item attributes or user-generated information, such as review or rating). Our methods, however, do not use any item-level features: our representation learning approach for the

existing-user scenario only utilizes the SKU of the three consumer goods in the source domain, and our conversion funnel transfer for the new-user scenario does not use any item-level features at all. Thus, our method is potentially more generalizable to other product domains.

Finally, in the new-user scenario, our approach differs from existing cross-domain recommender systems in that existing methods transfer users' product *preferences*, such as preferences for certain features (i.e., topics (Tang et al. 2012), aesthetic preference (Liu et al. 2020), etc.), or latent users' traits learned from various auxiliary sources and use the learned information as features in the target domain. Our proposed conversion behavior transfer, on the other hand, transfers users' behavior, which is the *mapping* from browsing behavior to final purchases learned from the source domain. Such behavioral similarity across product domains has not been studied or took advantage of by transfer learning to benefit the prediction task.

3. Conceptual Framework

We propose a conceptual framework to explain the rationale why users' online behavior associated with consumer goods, such as search and purchase, is helpful for predicting purchases of mutual funds. Recent findings suggest that users' traits can be inferred from their transaction and search patterns. On the other hand, another stream of literature also indicate that users' traits could significantly influence their financial decision-making, such as the investment on mutual funds. Therefore, we propose a framework as shown in Figure 1, which bridges the information obtained from search and purchase in consumer goods with the investment behavior in mutual funds.

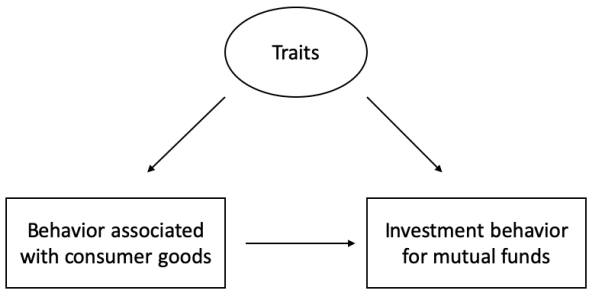

Figure 1 **Theoretical Framework Connecting Information in Consumer Goods with Mutual Funds Investment**

3.1. Purchase Behavior in Consumer Goods and Consumer Traits

Researchers have spent more than four decades studying whether the preference that users demonstrate from their purchase decision is stable across different product categories. In other words, is the preference in the purchase decision a consumer-specified trait? Literature in marketing (Blattberg et al. 1978, Ainslie and Rossi 1998, Inseong and Chintagunta 2007, Bell et al. 1999) shows that users' sensitivities to marketing mix variables such as price, advertising, and feature are stable across different categories. Hansen et al. (2006) indicates a significant correlation in households' preference for store brands. Singh et al. (2006) find that users' preferences for product attributes such as brand names and pack sizes are consistent across different product categories. Erdem et al. (2002) suggest that customers' sensitivities to loss and gain appear to be general shopping traits and are correlated across categories. These studies support the idea that certain unobservable household characteristics or "traits" influence purchase decisions regardless of product category (Ainslie and Rossi 1998, Singh et al. 2006, Hansen et al. 2006, Erdem et al. 2002).

Another stream of literature in psychology also provides evidence to support that traits are significantly correlated with purchase behavior. Gladstone et al. (2019) applies ensemble machine learning techniques to predict Big Five personality traits from transaction data. Their findings suggest that digital records of spending can be used to predict personality with modest and stable predictive accuracy. Matz et al. (2017) conducted field experiments with psychologically tailored advertising to reach millions of individuals. They find that the application of psychological targeting could significantly alter purchase behavior.

Besides purchase decision, literature also shows that traits can also be inferred from search processes such as variety seeking (Sharma et al. 2010), patterns of behavior collected with smartphones (Stachl et al. 2020), e.g., day- and night-time activity, and the preference over mobile vs. computer usage (Butt and Phillips 2008). All these pieces of the literature suggest that traits are important factors determining users' behavior in consumer goods, and they can be inferred from users' purchase and search history.

Accordingly, in our study, we generate features from users' prior purchase decisions in the consumer goods domain, such as the number of orders in different price levels, the number of clicks by different brands. We also create features based on users' search process, such as search effort, search channel, search timing, and search variety via the clickstream data. Details regarding the construction of these features are provided in Section 4.

3.2. Trait and Investment Decision

Users' traits have been found to influence investment decisions. Previous works (Durand et al. 2008, 2013, Oehler et al. 2018) show that investment decisions of individual investors, such as trading frequency and the way to seek information of the market, have a statistically significant association with personality, e.g., Big Five personality. Nicholson et al. (2005) suggests that personality profiles can be used to predict the propensity of risk-taking in the financial market. On the other hand, traits such as risk aversion and impatience, which are important factors driving investment decision, are also found to be related to users' choices in consumer goods. For example, users who are risk appetite are more likely to purchase generic brands than national brands (Erdem and Swait 2004, Khan et al. 2013). Kahneman and Tversky (1979) have established and developed the behavioral finance theory, in which the investors' characteristics are considered to be potentially strong predictors of investment decision. Pompian and Longo (2004) find different personal characteristics, such as preference of investment type, choice of information channel, will cause striking differences on individual investment decisions. Lan et al. (2017) study the predictability of investment behavior based on investors' personal characteristics of individual in China, and show that the income level and profit seeking desire have the most significant predictability on all types of investment behavior.

These results combined suggest that traits could influence both decisions of purchasing consumer goods and financial products. In practice, traits can be inferred from the transaction and search data in the category of consumer goods, and they could be used to predict the decision-making in financial products such as mutual funds. The theoretical framework we proposed in Figure 1 provides a rationale to support why the information extracted from purchase and search behavior in consumer goods can predict the investment decision in mutual funds, which also demonstrates that the prediction results we obtained from different methodologies are not arbitrary but with the theory backup.

4. Data Description and Problem Formulation

In this section, we provide a detailed description of our data and extracted features.

4.1. Data Overview

We obtained data from one of the largest shopping platforms in China. For our study we choose three categories of consumer goods, i.e. shampoo, toothpaste, and washer, which are

the most representative categories covering both non-durable (shampoo and toothpaste) and durable (washers) products. The dataset contains 10,000 randomly selected customers who have searched for mutual funds from May to December 2016 and have viewed at least one of shampoo, washer, and toothpaste on the same shopping platform from December 2014 to December 2016. The dataset includes the following information: 1) order history in mutual funds and three consumer goods categories for each user with product ID and time of order; 2) clickstream data in mutual funds and three consumer goods categories for each user, where each click is represented by product ID, device used, and time of click; 3) features of products in consumer goods including price and brand; 4) features of mutual funds including daily, weekly, and monthly return, risk level, and fees; 5) users' profile features, including education background, income level, risk averse level, and profit seeking level.

To provide a better understanding of the data, we show the summary statistics in Table 1 [1]. Our first goal in this paper is to investigate whether knowledge from users' shopping and browsing behavior in the consumer goods domain can be transferred to the financial products domain to improve the predictive accuracy of users' purchases of mutual funds.

Table 1 Summary of Statistics from December 2014 to December 2016

Summary of Statistics			Shampoo	Toothpaste	Washer	Mutual Fund
Number of clicks per user	App & Web	Mean	44.26	32.33	55.07	162.82
		Median	21.00	15.00	27.00	12.00
	App	Mean	16.18	13.73	36.62	163.34
		Median	7.00	6.00	12.00	12.00
	Web	Mean	42.15	31.14	40.72	17.92
		Median	18.00	14.00	19.00	4.00
Number of orders per user		Mean	4.26	4.86	1.43	18.40
		Median	3.00	3.00	1.00	11.00
Conversion rate per pser		Mean	0.37	0.53	0.14	3.48
		Median	0.15	0.23	0.05	1.00
Number of item			2,699	1,449	1,118	11,522
Number of item browsed / purchased at least once			2,491	1,445	1,033	6,728

[1] we remove 164 abnormal users which have one of the following conditions: 1) more than 5 clicks happen within one second; 2) clicked more than 1.5 times interquartile range away from the median within 24 hours).

4.2. Predictive Task: Purchase Intention Prediction

We first define the predictive task in our study: predicting users' purchase intention. Specifically, we aim to predict the number of orders a user will place within 24 hours of clicking a mutual fund product. Purchase intention prediction plays an important role in e-commerce to improve consumer experience and provide personalized recommendations (Esmeli et al. 2020, Schlosser et al. 2006). Such predictions, especially real-time predictions, enable platforms to take actions accordingly to improve purchase conversion rates (Sakar et al. 2019, Awad and Khalil 2012).

To solve this problem, we define two types of windows on mutual funds browsing data: *observation window* and *forecast window*, which are consecutive two time frames. The observation window covers a period of 7 days, going backward from a user's click of a mutual fund. A forecast window immediately follows an observation window and covers 24 hours period[2]. Each click, therefore, indicates the end of the observation window and the beginning of the forecast window, and thus activates a real-time prediction. See Figure 2 below for the explanation of two types of time window sessions. The target variable is defined as the number of orders placed within a forecast window.

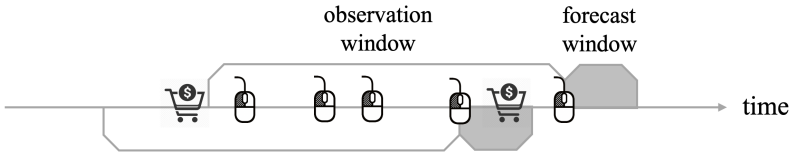

Figure 2 **An illustration of observation windows and forecast windows**

Baseline Features Construction We construct three sets of features as baseline features for the predictive models. The first set of features are mutual funds related features that are extracted from the observation window. We constructed seven features via feature engineering from the click stream data of mutual funds within the observation window. These features are listed in Table 8 in the Appendix. Note that similar features could be extracted for consumer goods, so we call them *common features*. We detail the theories supporting the extracted feature in Section 5. The second set of features are exclusive to

[2] We set the forecast window length to 24 hours after each click because more than 90% of orders were placed within the next 24 hours after a click, i.e., the time between 90% of the purchases and the last clicks before the purchases is less than 24 hours

mutual funds, so we call them *unique mutual funds features*. This set includes number of click from different risk levels, number of clicks based on monthly / annual / cumulative return rate. Those features are extracted from the observation window. These features are unique to mutual funds. The last set of features are from users' profiles listed in Table 9. These three sets of features are defined as baseline features for our predictive models. Note that baseline features do not include consumer goods features from Table 8.

Baseline Performance For this target problem (prediction of the purchase intention), there are a total of 219,542 clicks in this dataset, and therefore, 219,542 instances. We split the 219,542 instances by time: the first 80% for training, next 10% for validation, and the last 10% for testing. We apply a state-of-the-art machine learning model, deep learning, to the data[3]. R^2 is 53.68% evaluated on the test set. This performance is the benchmark for comparison later in the paper.

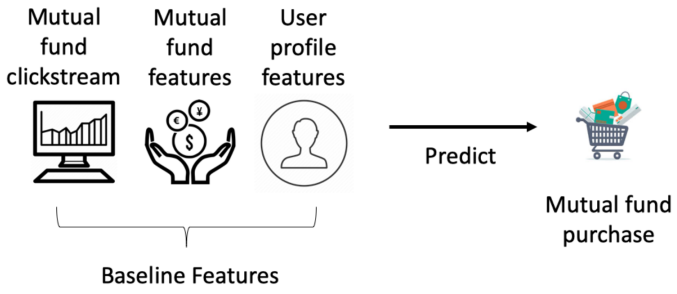

Figure 3 **Predicting mutual fund purchase intention using only baseline features.**

5. Knowledge Transfer for Existing Users

We first investigate the existing-user scenario, where we have target users' prior browsing and purchase history of consumer goods. Our goal is to *extract knowledge* from the consumer goods data, and concatenate the features with baseline features from Section 4.2 to evaluate whether the additional information from consumer goods domain can improve the predictive performance. Therefore, the key problem for this approach is how to extract knowledge from users' browsing history of consumer goods. See Figure 4 for an illustration of the knowledge transfer for existing users.

We test knowledge extraction via two approaches, a more manual approach utilizing theory-based feature engineering and user representation learning via machine learning.

[3] We also applied other machine learning models, such as random forest and SVM, but they did not perform as good as deep learning techniques. In addition, since two other models we will propose in the paper are both DNN-based. For better comparison, we will stick to DNN methods when comparing the predictivity of features

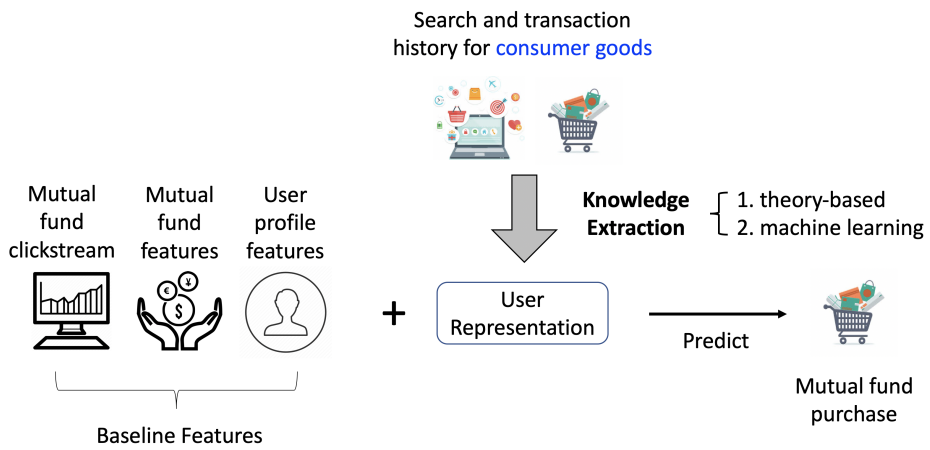

Figure 4 **Knowledge transfer for existing users via extracted features from consumer goods.**

5.1. Knowledge Extraction: Theory-Based User Representation

As discussed in Section 3, we generate features which could infer users' traits from their search and purchase data. We first generate user features based on theories from the literature. Below, we elaborate the theoretical support for our feature engineering.

5.1.1. Search Information The clickstream data records the details of customers' search path. For each consumer, records include consumer ID, the item that the consumer browsed, the time when the consumer browsed specific items, and the click channel (i.e., Mobile or Desktop). We describe how we extract search information to generate features of search effort, search timing, search channel and search variety.

Search effort We generate the feature of search effort by calculating the total number of clicks on the path to purchase. Previous literature (Tan and Tang 2013, Halder et al. 2010) have shown that users' personality influences the way they search information. Darley (1999) demonstrates that the drivers of search effort, i.e., time spent on acquiring product information, are different in terms of users' levels of esteem and self-esteem. We use the number of clicks to represent the effort that users spend on the search.

Search Channel The clickstream data in our study record users' search behavior in two different channels: mobile vs. desktop. We construct features by counting the number of clicks in two channels separately during the search process. Previous works indicate that users in mobile and desktop channels are different in terms of demographics, motivation, and personality (Butt and Phillips 2008). Therefore, we use these two features to represent users' preferences over the channels.

Search Timing Taking advantage of the time stamp in the clickstream data, we create features of the number of clicks that users have in different periods. Specifically, we divide one day into 4 time slots, i.e. 12 a.m - 6 a.m., 6 a.m. - 12 p.m., 12 p.m. - 6 p.m., 6 p.m. - 12 a.m, and count the number of clicks in each time slot. Stachl et al. (2020) show that users who are active in day time have distinctive personalities comparing with those who are active in the nighttime. We construct the features in terms of search timing as a proxy of users' active time.

Search Variety Our clickstream data observes the brand of the product in each click. We generate features of the number of clicks by different brands during the search process. Many studies in marketing and psychology have suggested that personality is an important determiner of users' variety seeking behavior. Sharma et al. (2010) conclude the findings from a survey that optimum stimulation level and self-monitoring level are strongly associated with variety-seeking motivation. Fernandes and Mandel (2014) explore the relationship between political ideology and variety seeking and demonstrate that conservatism is positively related to variety seeking. Thus, we generate features to represent the level of variety seeking.

5.1.2. Transaction information Our order data includes order time, item brand, and item price. In the following, we describe how we extract transaction information to generate features of frequency of historical orders and distribution of purchase price.

Consumer Loyalty We generate the feature of frequency of historical orders by counting the number of orders for each customer during the observation window. This feature represents how often users use the platform and their loyalty to the platform.

Purchasing Power To construct the features in terms of the purchase price, we create a categorical measure based on the quantiles of product transaction price in a given product category. We then count the number of orders with the product price in different levels. Previous literature show that purchase power is associated with income, education, socioeconomic status, etc., (Winkleby et al. 1992), which in turn have been shown to be correlated with users' traits (Judge et al. 1999, Bucciol et al. 2015, Andrei et al. 2015). In this study, we count the number of orders with the purchase price in different levels and use these features to represent the level of purchase power for each consumer.

We keep the feature construction in a simple way and it does not involve any item-level or product-level information, which would allow us to generalize similar construction to

other product categories. Table 8 (column *Consumer Goods*) in the Appendix lists features constructed for each of the three non-financial products [4], which include common features that can also be constructed for consumer goods and unique consumer goods features. Features in the columns *Consumer Goods* will later be combined with baseline features to examine whether such information improves the predictive accuracy.

5.2. Knowledge Extraction: Machine Learned User Representation

The second approach constructs user representations via machine learning that utilizes the representation learning and embedding techniques to obtain latent vectors for each user from their browsing history. The idea is to view users as different sets of products they have viewed and then learn a representation for each user by trying to reconstruct their sets of products, following the core idea in representation learning. However, there exists over five thousands unique products that users have clicked, which will lead to inefficient feature representation and curse of dimensions if simply using one-hot encoding to represent the products. Thus, we first apply an embedding technique to obtain a shorter and denser vector representation for items and then use representation learning to learn a latent vector for each user from the set of items s/he has browsed.

Product Embedding via Item2Vec The first step is to learn a vector q to represent a product, each represented by a SKU. We apply the item2vec technique (Oren and Noam 2016) which exploits the co-occurrence of products in users' click streams. This is similar to word embedding (Mikolov et al. 2013), which utilizes the co-occurrences of words in a moving window to learn a vector to represent each word. Here we embed the item SKU. We define a moving window of 30 minutes and only examine products of the same category when applying the embedding model. Thus, toothpaste, shampoo, and washers are embedded into different feature spaces, which is intended since across products the features are not comparable. The item2vec method will return an embedded vector $q_{k,j}^{(i)}$ for each item, where i is the user index, product k is the product category (shampoo, toothpaste or washer) and j is the item index.

User Representation Learning Let $\mathcal{U}$ represent the set of users from data. User i is denoted as $u^{(i)} \in \mathcal{U}$, which is a one-hot encoding representation with zero everywhere except the $i-$the element being one. Since there are a total of 10,000 users, $u^{(i)}$ is a vector of length

[4] Note that the four products' features do not completely align.

10,000. Let $C^{(i)}$ represent the three categories of products that user i has purchased and $C^{(i)} \subset \{1, 2, 3\}$. $C^{(i)}$ is different across users. Let $Z_k^{(i)}$ represent the set of items of in category k user i has clicked.

Our goal is to learn a denser and much shorter embedding vector $\hat{u}^{(i)}$ for each user $u^{(i)}$. Each user has browsed or purchased multiple categories of products. We use three transformation matrices W_k to predict from a user embedding $\hat{u}^{(i)}$ to his corresponding products of type $C^{(i)}$. This design can be generalized to more products in a real setting.

$$\hat{u}^{(i)} = V u^{(i)} \tag{1}$$

$$\hat{q}_k^{(i)} = W_k \hat{u}^{(i)}, \forall k \in \{1, 2, 3\} \tag{2}$$

V is an embedding layer that maps each one-hot-encoded vector to the corresponding embedded representation. W_k represents the projection from a user to his clicked products of category k. We set the length of $\hat{u}_i$ to be 10, after experimenting with sizes from 2 to 15.

To learn the embedding, we use standard reconstruction error (Ng et al. 2011), represented by the difference between the true true vector $q_{k,j}^{(i)}$ and the reconstructed vector $\hat{q}_k^{(i)}$. The design of the objective is motivated by representation learning that a good embedding vector should be able to predict the product a user will click. The loss function is defined as the following:

$$\min \sum_{u^{(i)} \in \mathcal{U}} \sum_{k \in C^{(i)}} \sum_{j \in \{1, \cdots, |Z_k^{(i)}|\}} (\hat{q}_k^{(i)} - q_{k,j}^{(i)})^2 \tag{3}$$

Here, $q_{k,j}^{(i)}$ is the true features of an item, indexed by j, for a product k and $\hat{q}_k^{(i)}$ is the corresponding reconstructed feature vector for $q_{k,j}^{(i)}$.

The benefit of this approach is that the embedding representation is fully data-driven and optimized on click stream data, which is different from feature engineering with the help of human knowledge for feature extraction.

5.3. Predictive Performance Evaluation

In this section, we evaluate the two transfer approaches and compare their performance with the baseline from Section 4.2. We partition the data by users' clicks by time: the first 80% of users' clicks are used for training, next 10% for validation, and the remaining 10% for testing. We do that for each user separately to make sure all users appear in training,

Table 2 Performance of Models Using User Representations Extracted from Consumer Goods in the Existing-User Scenario

Features	**Theory-Based**		**Machine Learned**	
	R^2	Lift	R^2	Lift
Baseline + Shampoo	60.25%	6.57%	62.30%	8.68%
Baseline + Toothpaste	59.47%	5.79%	63.09%	9.41%
Baseline + Washer	60.32%	6.64%	63.53%	9.85%
Baseline + All Consumer Goods	62.65%	8.97%	66.09%	12.41%

Note: Baseline Performance is 53.68%.

validation, and test sets. For both user representation approaches discussed previously, we implement four models, using data from shampoo, toothpaste, and washer, separately and combined. Each user representation is then concatenated with the baseline features and fed into a fully connected neural network. We use the validation set to tune the number of layers and number of nodes. We then report the performance on the test set in Table 2, which includes the R^2 and *lift*, where the lift is the improvement in R^2 compared to the baseline performance of 53.68%.

Results show that users' information extracted from consumer goods via both methods significantly improved the prediction performance in the mutual funds domain. Meanwhile, using information from more consumer goods achieves bigger improvements. The theory-based feature engineering approach is able to improve the performance to up to 8.97% by using information from all three consumer goods. User representations learned via machine learning can achieve up to 66.09% in R^2, an 12.41% improvement from the baseline.

Both methods extract features from consumer goods that characterize users' product preferences or traits. The main difference between theory-based and machine learning approaches is the way for extracting features. Theory-based manual feature engineering relies on human intuition and theory from the literature while machine learned user representation is completely data-driven, relying on state-of-the-art deep learning techniques, which turns out to be more powerful in learning latent traits of users.

6. Knowledge Transfer for New Users

The methods above only apply to existing users, since they need to learn a user representation for the target users from a source domain. For new users, however, such information does not exist as they have no prior activities in existing domains. Therefore, in this section, we propose a new solution that transfers the conversion funnel from consumer goods to financial products. Conversion funnel describes the journey a consumer takes through search system and finally converting to a sale. Literature (De Haan et al. 2016) shows that users' behavior within the conversion funnel such as the response to advertisement exhibit consistency across product categories. We hypothesize that conversion funnel in consumer goods shares similarities with that in financial products, i.e., the mapping from the browsing behavior (input) to orders (output) for toothpaste is similar to the input-output mapping for mutual funds across users. See Figure 5 for an illustration of our idea. With this conjecture, we expect such mapping learned in consumer goods could improve the prediction performance of purchases of mutual funds.

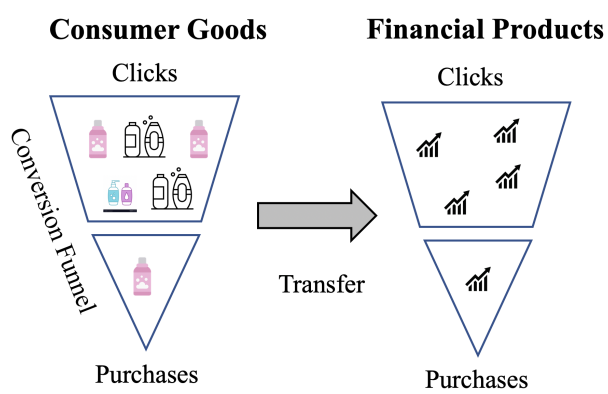

Figure 5 **Knowledge transfer via Conversion Funnel.**

6.1. Conversion Funnel Transfer

To formally test our conjecture, we first develop the model of how to learn the conversion funnel. We first discuss how to transfer the conversion funnel (i.e., mapping) for single product categories and then discuss how we can aggregate the mappings across multiple products.

Conversion Funnel Transfer from a Single Product Category First, we build a neural network to predict the number of orders of product k ($k = 1, 2, 3$ represent toothpaste, shampoo, and washer, respectively) following the same baseline framework described in Section 4.2, to align with our focal tasks of predicting the number of orders of mutual funds. This time, the observation window, forecast window as well as the target variable are defined on product k. To make the features align across products, we only use the *common features* listed in Table 8, such that for all the four types of products, the features involved in this method have the same dimension and meaning, which makes a transfer possible later. We then train a three-layer fully connected neural network f_k on the training data of product k and $f_k(\cdot) := \Psi_k^3(\Psi_k^2(\Psi_k^1(\cdot)))$. Given $\mathbf{x}_k^{(i)}$, the input features for type k product, the output is expressed as

$$\hat{z}_k^{(i)} = \Psi_k^3(\Psi_k^2(\Psi_k^1(\mathbf{x}_k^{(i)})). \tag{4}$$

$\hat{z}_k^{(i)}$ is the predicted number of orders of type k product for user i in the forecast window. Ψ_k^h at layer h is Relu with dropout rate of 10%: $\Psi_k^h(\mathbf{x}_k^{(i)}) = \text{ReLu}(\mathbf{w}_k^h \text{Dropout}(\mathbf{x}_k^{(i)}) + b_k^h)$. We save the first two hidden layers, which will be transferred into mutual funds prediction model later. Please see Figure 6 for mutual funds prediction process from pre-trained model to individual product.

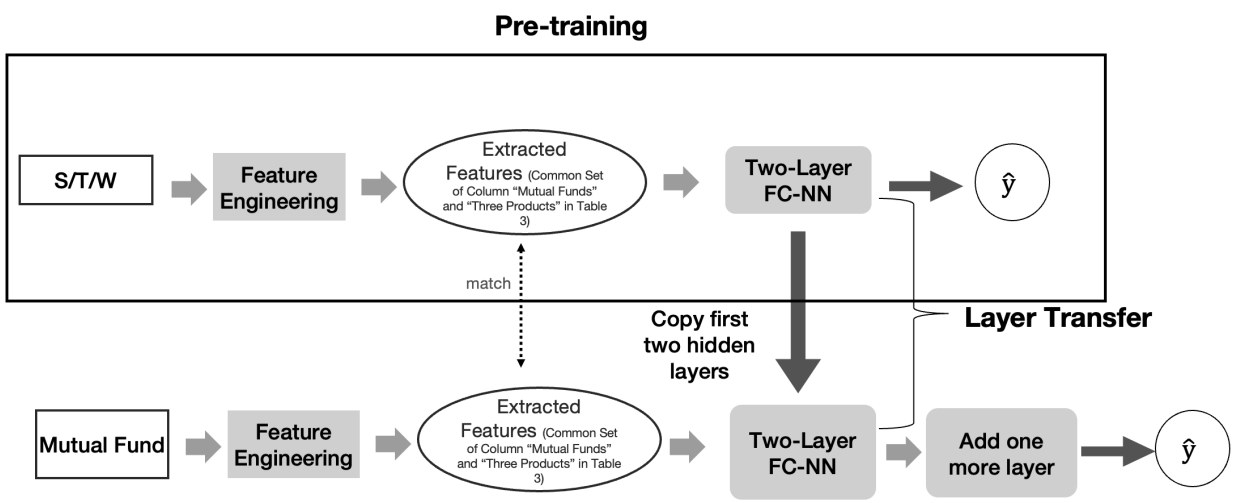

Figure 6 **Conversion Behavior Transfer**

After obtaining the model f_k, we train the model f_u on the mutual funds baseline features. f_u has the same architecture as f_k and the first two layers are directly copied

from f_k and set non-trainable. The remaining layer continues to train on the mutual funds data. In case of any under-fitting problems, one can add additional trainable layers.

$$\hat{y}_{\mathrm{m}}^{(i)} = \Psi_{\mathrm{m}}(\Psi_k^2(\Psi_k^1(\mathbf{x}_{\mathrm{m}}^{(i)})). \tag{5}$$

Here, coefficients $\mathbf{w}_k^1, \mathbf{w}_k^2, b_k^1, b_k^2$ in layers Ψ_k^1, Ψ_k^2 are copied from models (4). $\mathbf{x}_m^{(i)}$ represents feature that align with $\mathbf{x}_k^{(i)}$ but extracted from the browsing history for mutual funds. Then the model is trained by minimizing the MSE loss between $\hat{\mathbf{y}}_{\mathrm{m}}$ and $\mathbf{y}$. The layer transformation $\Psi_k^2(\Psi_k^1(\cdot))$ is part of the *conversion funnel* that is transferred across product domains.

Transfer via Conversion Behavior from All Product Categories Additionally, we propose a model, $f_{u,\mathrm{all}}$ that can utilize multiple products by concatenating the output of the second layer of f_k and feed it to a third layer.

$$r_k^{(i)} = \Psi_k^2(\Psi_k^1(\mathbf{x}_{\mathrm{m}}^{(i)}), \forall k \in \{1, 2, 3\} \tag{6}$$

$$\hat{y}_{\mathrm{m,all}}^{(i)} = \Psi_{\mathrm{mf,\ all}}([r_1^{(i)}, r_2^{(i)}, r_3^{(i)}]) \tag{7}$$

$\Psi_{\mathrm{mf,\ all}}$ is a layer that needs to be learned.

Conversion behavior transfer does not need to obtain a target user's prior shopping record, which is required for the user representation learning methods in the previous section. The users' conversion behavior will be captured and derived from other users' shopping records in the source domain, which are only required in the training stage for pre-trained model on individual product (i.e. shampoo, toothpaste, and washer). The learned conversion behavior is stored in weights from pre-trained model and will be transferred and aggregated later for predicting new users' purchase decision in mutual funds. In testing, the input is only the features related to mutual funds ($\mathbf{x}_{\mathrm{m}}^{(i)}$ represents the seven common features from Table 8) and no information from consumer goods domain is needed, so the model works for new users.

6.2. Predictive Performance Evaluation

We evaluate the performance of conversion behavior transfer on new users. We partition the data *by users*, instead of by transactions in the previous experiments. 80% of users are used for training, 10% of users are used for validation, and the remaining 10% of users are used for testing. These three sets are mutually exclusive because we need to ensure that

Table 3 Performance of Conversion Behavior Transfer on New Users

Features	**Conversion Behavior Transfer**	
	R^2	Lift
Baseline + Shampoo	39.34%	8.38%
Baseline + Toothpaste	38.88%	7.92%
Baseline + Washer	40.48%	9.52%
Baseline + All Consumer Goods	45.46%	14.50%

Note: Baseline Performance is 30.96%.

users selected for testing data are indeed "new users". In addition, the consumer goods data *only participate in training* and no such information is used for testing, to mimic the real-world scenario where new users do not have prior information in existing domains.

We first re-train the baseline model and report the performance in Table 3. Note that the performance drops to 30.96% compared to 53.68% from Table 2. This is because the predictive problem now becomes even more challenging since the model is trained on one group of users while making predictions to a totally different set of users. On the other hand, it is also more interesting to verify whether a machine learning guided "transfer" could work.

We build four models, three of them using one category of consumer goods, respectively, and the fourth one using all consumer goods. The experimental results evaluated on the test set are listed in Table 3. Results show that transferring the conversion funnel in consumer goods to mutual funds can significantly improve the predictive performance. Specifically, using the conversion funnels learned from all three categories of consumer goods achieves the largest improvement, 14.5% in R^2. These findings evidence the effectiveness of using transfer learning for addressing the cold start problem of new user scenario. Results also imply that the conversion funnel between the two domains bear some similarities, even if the two domains are entirely different from each other.

7. Heterogeneity Analysis

We study the effect of knowledge transfer on users with different features. Then we compare users' information learned from consumer goods with self-reported user profile features in terms of predictivity.

7.1. User Heterogeneity

We have shown so far that the information in consumer goods categories can improve the prediction performance of users' purchases of mutual funds. In this section, we aim to investigate how such performance lift varies across users with different characteristics. Specifically, we consider in four features in users' profile: education, income, and their investment preferences including profit goal and risk preference. Table 10 in the Appendix shows the description of these features. We examine the user heterogeneity only for the existing-user scenario since the prediction model in new-user scenario is not based on the same users.

We first apply clustering to partition users into groups with similar characteristics. Since the features are categorical, we use K-mode clustering (Chaturvedi et al. 2001) and use the elbow method to find the number of clusters to be two. The first group mainly consists of users that are well-educated with stable high income, more risk aggressive, and good at controlling family spending budget, which is defined as *high socioeconomic group*. The second cluster consists of users that are less educated, with no stable or less income, more risk averse, and usually spending over the family budget. It is defined as *low socioeconomic group*. The summary statistics of the features in the two groups are included in Table 10 in the Appendix. The description of the clusters is shown in Table 4. The characteristics of the two groups are consistent with the classic definitions of low and high socioeconomic groups from the literature (Vyas and Kumaranayake 2006).

Table 4 K-modes Clustering on Users' Socioeconomic statuses and Risk Preferences

Clustering Group	Size	Description
High Socioeconomic Group	26.7%	higher income, better educated, seeking more profit, higher risk tolerance
Low Socioeconomic Group	73.3%	lower income, under educated, seeking asset preservation, lower risk tolerance

We first re-evaluate the baseline performance for each cluster. Note that the baseline model is the same as the one in Table 2 and we re-computed the R^2 for each cluster. Results show that the baseline model performance in the low socioeconomic group is much higher than that in the high socioeconomic group, i.e., 45.23% vs. 18.72% in R^2.

We then re-evaluate the method using machine learned user representations for each cluster. The performance is reported in Table 5. Results show that the model for users of high

socioeconomic group can achieve 20.11% performance lift with machine learned user representation, while the model for users of low socioeconomic group gain a 12.54% performance increase. Thus, the performance gap between the two clusters is reduced from approximately 26% to 18% after including machine learned user representations. One potential reason for the performance difference in two clustered groups is users' behavior heterogeneity. Users in high socioeconomic group tend to have more options when making investment decision due to their stronger economic capabilities, while the low socioeconomic group is more constrained. Therefore, the model performance for users in the low socioeconomic group is generally much higher than users in the high socioeconomic group, as behaviors of low socioeconomic group are more homogeneous and predictable. We also observe that information extracted from consumer goods can improve the model performance more for users in the high socioeconomic group. Since behaviors of high socioeconomic users are more heterogeneous, user profile features are insufficient of characterizing their behavior for investment decisions. Our findings suggest that information extracted from users' activities in consumer goods categories can potentially capture users' traits better than using user profile features.

Table 5 Performance(R^2) on clustered group

Clustering Group	**Baseline**	**Machine Learned User Representation**	**Lift**
High socioeconomic Group	18.72%	38.83%	20.11%
Low socioeconomic Group	45.23%	57.77%	12.54%

7.2. Which Information is More Valuable? - Consumer Goods versus User Profile

Financial institutes used to rely heavily on self-reported user profile to make prediction of users' investment decision. The user profile describes users' characteristics including information of education, income, and investment preference. Our analyses above, on the other hand, have demonstrated that the knowledge extracted from users' behavior in consumer goods categories can also characterize users' traits, and significantly improve the performance of predicting users' purchases of mutual funds. In this section, we examine which

set of information is more valuable in predicting mutual funds purchases, self-reported user profile or user traits learned from their shopping activities in consumer goods.

To answer the question above, we evaluate two sets of features. The first one is the baseline features described in Section 4, which include mutual funds related features and user profiles. The other feature set includes mutual fund related features and machine learned user representation obtained from three consumer goods categories. We build fully connected neural network models with these two different sets of features respectively and report the performance in the first Column in Table 6. Results show that the model with user representation learned from consumer goods have statistically higher prediction performance than the one with user profile features. We then examine the model performance in different user groups similar to subsection 7.1. The second and third Columns in Table 6 show the model performance with the different sets of features for users in the low socioeconomic group and the high socioeconomic group, respectively. Results demonstrate that user representations learned from consumer goods can achieve better model performance in both groups. It implies that user behavior can better signal users' purchase intention than fixed user profile features.

Table 6 Performance (R^2) of Models Using Different Information

User Group	All Users	Low Socioeconomic Group	High Socioeconomic Group
Mutual Funds + User Profile	53.68%	42.79%	11.38%
Mutual Funds + All Consumer Goods	62.65%	54.98%	34.75%

8. Discussion and Robustness Check

This study demonstrates the value of knowledge about users learned from consumer goods to predictions in the financial products domain. Using proposed machine learning methods, we are able to achieve performance lift using information extracted from users' prior activities associated with consumer goods to predict their purchases of mutual funds. The rationale for such performance lift is supported by literature on consumer behavior and psychology. Below we first reflect on why the proposed methods are viable in our context and then perform a robustness check where we reformulate the prediction problem to a classification problem.

8.1. Reflection on Why Representation Learning and Transfer Learning Work

Our use of machine learning techniques is based on the close connections between how the algorithms work and the theoretical foundations on the relations between user behaviors across different products. The critical components in our solutions are representation learning in the existing-user scenario and transfer learning in the new-user scenario. We discuss their connections to the theoretical foundations.

The theoretical framework in Section 3 suggests that consumers' behavior is a result of their internal traits, which influence their shopping behavior in different product domains, including consumer goods and financial products. Therefore, an important step in the existing-user scenario is to obtain a representation of traits for each user from the consumer goods domain, such that this information can be used in the financial products domain. Traits can be represented by demographic features such as gender, age, etc, or manually constructed features with actual meanings, such as search effort, search depth, etc, based on related theories. The latter is adopted by our theory-based approach in the existing-user scenario. In addition, traits could also be reflected by latent representations learned from users' browsing activities. The fundamental idea in the machine learning approach is to view each user as represented by the items they have clicked. Then, the learning objective for the representation learning is to reconstruct the viewed items for each user to obtain a concise vector representation. One can think of representation learning in this context as a method to obtain a vectorized representation for a user from the original set of clicked items. Unlike feature engineering that directly constructs features, vectors obtained via representation learning are completely data-driven for maximally preserving users' clicks information in the original data. In summary, for the existing-user scenario, users' traits are learned via representation learning and transferred.

While representation learning is used based on the assumption that users' traits are reflected by clicks, our second approach for the new-user scenario, conversion funnel transfer, is developed based on the similarity of conversion behavior in different product domains. The conversion funnel can be effectively represented as the *mapping* from the browsing history to the purchases, which can be represented by a model processing the input (browsing history) to get a prediction (purchases). Therefore, when there is a similarity in the mapping (conversion funnel), as proved by the literature (De Haan et al. 2016) on some product categories, one can re-use the model to obtain a prediction from the

input in a different product domain. This process resembles the idea of transfer learning, which takes an existing model and continues training with new data. In other words, in the new-user scenario, users' conversion behavior is transferred.

In conclusion, the two machine learning techniques bear a remarkable resemblance to theories from prior literature on consumer behavior and psychology.

8.2. Robustness Check

In order to further validate the conclusions so far, we perform a robustness check by redefining our predictive task from regression to classification. Instead of predicting the number of orders, the target variable becomes a binary variable, i.e. whether the user will place orders for the mutual funds within the forecast window. We implement the same three approaches with 5 feature combinations (including baseline features). Since the data is unbalanced with about 21% positive predictions, we use AUC to measure the classification performance and quantify the value of information from different products. See Table 7. According to the results, we observe consistent performance lifts in AUC from the classification outcomes in comparison to our previous regression task. This means, regardless of the predictive task formulation, transferring knowledge from consumer goods domain to the mutual funds domain is beneficial to users' purchase intention prediction and our methods remain effective.

Table 7 Performance (AUC) of Classification Models Using Different Transfer Approaches

Features	**Existing User**				**New User**	
	Theory Based User Representation		Machine Learned User Representation		Conversion Behavior Transfer	
	AUC	Lift	AUC	Lift	AUC	Lift
Baseline + Shampoo	0.715	0.014	0.769	0.068	0.612	0.089
Baseline + Toothpaste	0.711	0.010	0.780	0.079	0.614	0.091
Baseline + Washer	0.714	0.013	0.752	0.051	0.606	0.083
Baseline + All Consumer Goods	0.743	0.042	0.806	0.105	0.641	0.118

Note: Baseline Performance is 0.701 for existing user and 0.523 for new user

9. Conclusion

This study demonstrates the value of exploiting knowledge learned from the consumer goods domain to benefit customer purchase behavior predictions in a new financial product domain. The machine learning solutions are effective in both existing-user and new-user

scenarios. The value of the "transferred" information has also been further explored for users with different socioeconomic statuses and risk preferences. In addition, this knowledge learned from users' behaviors in the consumer goods domain has also proved to be more predictive than users' profile features.

9.1. Theoretical and Method Contribution

The literature in finance used to study individual investors' behavior from the angles of individual characteristics such as gender and age (Cronqvist et al. 2011, Sicherman et al. 2016, Cronqvist and Siegel 2010), culture (Grinblatt and Keloharju 2001) and historical investment behavior (Malmendier 2021, Barnea et al. 2010). The previous research on purchase intention prediction for mutual funds and modeling is limited because investment decisions depend on several external factors, such as psychological and emotional activities (Boda 2018), which are typically immeasurable. The existing literature is also limited regarding whether and how users' behavior on consumer goods can be powerful predictors for their purchases forecasting, while our study fills up this research gap. Our paper provides a new aspect in understanding and predicting investors' behavior. Our findings suggest that mutual funds purchase decisions can be traced from the purchase and search behavior in the consumer goods domain. This paper is seeking to open a new research subject where scholars could better understand investors' behavior from their activities in other areas such as consumer goods purchase, patterns of leisure time spending, etc.

Our study also contributes to the literature on cross domain predictions in two aspects. First, we show that using machine learning techniques can effectively transfer knowledge from two seemingly unrelated product domains. While there has been extensive research on cross-domain recommendations, they have mainly been investigating relevant domains where they can take advantage of content-relevance or item-relevance. Our paper studies the transfer between consumer goods and financial products, for the first time, where there are no above relevance, and we prove that the transfer is still possible and beneficial. Second, while most existing cross-domain recommender systems transfer users' product preferences or traits from the source domain, our solution to the new-user scenario proposes to transfer users' behavior (conversion funnels in our context) from the source domain. Our solution is inspired by existing literature that the conversion funnel is similar on some product categories. We extend it to a broader context, product domains. The success of the transfer may open up interesting research opportunities for studying the similarity in users' browsing, searching, and shopping behavior across different product domains.

9.2. Managerial Implication

This study has several implications for research on data valuation and product expansion.

First, user behavior data can achieve larger lifts than user profiles (e.g., age, gender, etc.) in predicting their purchases of mutual funds. Our study suggests that the financial industry can become more accurate in predicting users' decisions with the help of the information in non-financial areas, such as their transaction history in consumer goods. In addition, the findings from our heterogeneity analyses indicate that high income and well-educated clients tend to have higher personalized service improvement with such information. Our findings bring valuable insights to the financial industry, especially for those subsidiary financial companies (e.g., Ant Group, JD Digits) established from their parent e-commerce companies (e.g., Alibaba, JD.com). There is a huge privilege for these companies that plan to expand a new financial product because they can make significant use of users' prior shopping information on existing categories of products.

Second, in addition to proving the value of data, our paper also proves the value of methods. The better performance of machine learned representations suggests that advanced machine learning methods are better solutions for learning and exploiting useful information from data than theory-based manual feature engineering. As e-commerce platforms are gathering user data with an increasing volume and complexity, it is critical for the platforms to adopt appropriate and power machine learning approaches for extracting information from these data.

Finally, our findings can be generalized beyond the financial industry. Many e-commerce companies are trying to expand their product domains in recent years. For example, Amazon has introduced Amazon Pharmacy in 2020 to help customers conveniently purchase their prescription medications (Amazon 2020). Taobao, the largest e-commerce platform in Asia, has established Feizhu Travel that focuses on consumer and business travel products. They face the common challenge at the early stage of their launch, of not having enough data in the new domain to enable reliable predictions of users' purchase decisions. Thus, these e-commerce platforms would not be able to provide a series of online services such as personalized advertisement, etc. Our study makes a simple yet effective proposal of using users' shopping data from existing domains on newly expanded product domain and our proposed methods are generalizable to different product domains since, unlike many other cross-domain recommender systems, our methods do not require detailed item or content information.

References

Ainslie A, Rossi PE (1998) Similarities in choice behavior across product categories. *Marketing Science* 17(2):91–106, URL http://dx.doi.org/10.1287/mksc.17.2.91.

Amazon (2020) Introducing amazon pharmacy: Prescription medications delivered. URL https://press.aboutamazon.com/news-releases/news-release-details/introducing-amazon-pharmacy-prescription-medications-delivered.

Andrei F, Mancini G, Mazzoni E, Russo P, Baldaro B (2015) Social status and its link with personality dimensions, trait emotional intelligence, and scholastic achievement in children and early adolescents. *Learning and Individual Differences* 42:97–105.

Awad MA, Khalil I (2012) Prediction of user's web-browsing behavior: Application of markov model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42(4):1131–1142.

Barnea A, Cronqvist H, Siegel S (2010) Nature or nurture: What determines investor behavior? *Journal of Financial Economics* 98:583–604, URL http://dx.doi.org/10.1016/j.jfineco.2010.08.001.

Bell D, Chiang J, Padmanabhan V (1999) The decomposition of promotional response: An empirical generalization. *Marketing Science* 18, URL http://dx.doi.org/10.1287/mksc.18.4.504.

Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35:1798–1828, URL http://dx.doi.org/10.1109/TPAMI.2013.50.

Bengio Y, Lamblin P, Popovici D, Larochelle H (2006) Greedy layer-wise training of deep networks. *Proceedings of the 19th International Conference on Neural Information Processing Systems*, 153–160, NIPS'06 (Cambridge, MA, USA: MIT Press).

Berkovsky S, Kuflik T, Ricci F (2007) Cross-domain mediation in collaborative filtering. *International Conference on User Modeling*, 355–359 (Springer).

Blattberg R, Buesing T, Peacock P, Sen S (1978) Identifying the deal prone segment. *Journal of Marketing Research* 15(3):369–377, URL http://dx.doi.org/10.1177/002224377801500307.

Boda JR (2018) Investor's psychology in investment decision making: A behavioral finance approach. *International Journal of Pure and Applied Mathematics* 119.

Bucciol A, Cavasso B, Zarri L (2015) Social status and personality traits. *Journal of Economic Psychology* 51:245–260.

Butt S, Phillips JG (2008) Personality and self reported mobile phone use. *Computers in Human Behavior* 24(2):346–360, ISSN 0747-5632, URL http://dx.doi.org/https://doi.org/10.1016/j.chb.2007.01.019, part Special Issue: Cognition and Exploratory Learning in Digital Age.

Cao B, Liu NN, Yang Q (2010) Transfer learning for collective link prediction in multiple heterogenous domains. *ICML*.

CBInsights (2022) Everything you need to know about what amazon is doing in financial services. *CBInsights* URL https://www.cbinsights.com/research/report/amazon-across-financial-services-fintech/.

Chaturvedi A, Green PE, Caroll JD (2001) K-modes clustering. *J. Classif.* 18(1):35–55, ISSN 0176-4268, URL http://dx.doi.org/10.1007/s00357-001-0004-3.

Connolly B (2021) Top amazon product categories. *JungleScout* URL https://www.junglescout.com/blog/amazon-product-categories/.

Cronqvist H, Makhija A, Yonker S, Professor R, Agarwal S, Ambrose B, Ben-David Z, Coles J, Deangelo H, Fahlenbrach R, Funder D, Halov N, Helland E, Hughson E, Karolyi A, Karpoff J, Klasa S, Kumar A, Loughran T, Hom (2011) Behavioral consistency in corporate finance: Ceo personal and corporate leverage .

Cronqvist H, Siegel S (2010) The origins of savings behavior. *Institute for Financial Research, SIFR Research Report Series* 123, URL http://dx.doi.org/10.2139/ssrn.1649790.

Darley WK (1999) The relationship of antecedents of search and self-esteem to adolescent search effort and perceived product knowledge. *Psychology & Marketing* 16(5):409–427.

De Haan E, Wiesel T, Pauwels K (2016) The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing* 33(3):491–507.

de Souza Pereira Moreira G, Ferreira F, da Cunha AM (2018) News session-based recommendations using deep neural networks. *Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems*, 15–23.

Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. *Journal of the American society for information science* 41(6):391–407.

Durand RB, Newby R, Sanghani J (2008) An intimate portrait of the individual investor. *Journal of Behavioral Finance* 9(4):193–208, URL http://dx.doi.org/10.1080/15427560802341020.

Durand V, Hieneman M, Clarke S, Wang M, Rinaldi M (2013) Positive family intervention for severe challenging behavior i a multisite randomized clinical trial. *Journal of Positive Behavior Interventions* 15:133–143, URL http://dx.doi.org/10.1177/1098300712458324.

Elkahky AM, Song Y, He X (2015) A multi-view deep learning approach for cross domain user modeling in recommendation systems. *Proceedings of the 24th international conference on world wide web*, 278–288.

Erdem T, Swait J (2004) Brand credibility, brand consideration, and choice. *Journal of Consumer Research - J CONSUM RES* 31:191–198, URL http://dx.doi.org/10.1086/383434.

Erdem T, Swait J, Louviere J (2002) The impact of brand credibility on consumer price sensitivity. *International Journal of Research in Marketing* 19:1–19, URL http://dx.doi.org/10.1016/S0167-8116(01)00048-9.

Esmeli R, Bader-El-Den M, Abdullahi H (2020) Towards early purchase intention prediction in online session based retailing systems. *Electronic Markets* 1–19.

Farseev A, Samborskii I, Filchenkov A, Chua TS (2017) Cross-domain recommendation via clustering on multi-layer graphs. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 195–204.

Fernandes D, Mandel N (2014) Political conservatism and variety-seeking. *Journal of Consumer Psychology* 24(1):79–86, ISSN 1057-7408, URL `http://dx.doi.org/https://doi.org/10.1016/j.jcps.2013.05.003`.

Fernández-Tobías I, Cantador I (2014) Exploiting social tags in matrix factorization models for cross-domain collaborative filtering. *CBRecSys@ RecSys*, 34–41 (Citeseer).

Fu W, Peng Z, Wang S, Xu Y, Li J (2019) Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 94–101.

Gao C, Chen X, Feng F, Zhao K, He X, Li Y, Jin D (2019) Cross-domain recommendation without sharing user-relevant data. *The world wide web conference*, 491–502.

Gao S, Luo H, Chen D, Li S, Gallinari P, Guo J (2013) Cross-domain recommendation via cluster-level latent factor model. *Joint European conference on machine learning and knowledge discovery in databases*, 161–176 (Springer).

Gladstone J, Matz S, Lemaire A (2019) Can psychological traits be inferred from spending? evidence from transaction data. *Psychological Science* 30:095679761984943, URL `http://dx.doi.org/10.1177/0956797619849435`.

Goodfellow I, Bengio Y, Courville A (2016) *Deep learning* (MIT press).

Grbovic M, Cheng H (2018) Real-time personalization using embeddings for search ranking at airbnb. *SIGKDD*, 311–320, KDD '18.

Grbovic M, Djuric N, Radosavljevic V, Silvestri F, Bhamidipati N (2015) Context- and content-aware embeddings for query rewriting in sponsored search. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 383–392, SIGIR '15 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450336215, URL `http://dx.doi.org/10.1145/2766462.2767709`.

Grinblatt M, Keloharju M (2001) What makes investors trade? *The Journal of Finance* 56(2):589–616.

Hafshejani ZY, Kaedi M, Fatemi A (2018) Improving sparsity and new user problems in collaborative filtering by clustering the personality factors. *Electronic Commerce Research* 18(4):813–836.

Halder S, Ray A, Chakrabarty P (2010) Gender differences in information seeking behavior in three universities in west bengal, india. *The International Information & Library Review* 42(4):242–251, ISSN 1057-2317, URL `http://dx.doi.org/https://doi.org/10.1016/j.iilr.2010.10.004`.

Hansen K, Singh V, Chintagunta P (2006) Understanding store-brand purchase behavior across categories. *Marketing Science* 25:75–90, URL http://dx.doi.org/10.1287/mksc.1050.0151.

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

He X, Liao L, Zhang H, Nie L, Hu X, Chua TS (2017) Neural collaborative filtering. *Proceedings of the 26th international conference on world wide web*, 173–182.

Hinton G, Deng L, Yu D, Dahl GE, Mohamed Ar, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine* 29(6):82–97, URL http://dx.doi.org/10.1109/MSP.2012.2205597.

Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput.* 18(7):1527–1554, ISSN 0899-7667, URL http://dx.doi.org/10.1162/neco.2006.18.7.1527.

Inseong S, Chintagunta P (2007) A discrete–continuous model for multicategory purchase behavior of households. *Journal of Marketing Research - J MARKET RES-CHICAGO* 44:595–612, URL http://dx.doi.org/10.1509/jmkr.44.4.595.

Jansen BJ, Schuster S (2011) Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research* 12(1):1.

Jenkins P (2019) Clickgraph: Web page embedding using clickstream data for multitask learning. 37–41, WWW '19 (New York, NY, USA), ISBN 9781450366755, URL http://dx.doi.org/10.1145/3308560.3314198.

Judge TA, Higgins CA, Thoresen CJ, Barrick MR (1999) The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology* 52(3):621–652, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1744-6570.1999.tb00174.x.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291, ISSN 00129682, 14680262, URL http://www.jstor.org/stable/1914185.

Kanagawa H, Kobayashi H, Shimizu N, Tagami Y, Suzuki T (2019) Cross-domain recommendation via deep domain adaptation. *European Conference on Information Retrieval*, 20–29 (Springer).

Kang S, Hwang J, Lee D, Yu H (2019) Semi-supervised learning for cross-domain recommendation to cold-start users. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1563–1572.

Khan R, Misra K, Singh V (2013) Ideology and brand consumption. *Psychological science* 24, URL http://dx.doi.org/10.1177/0956797612457379.

Kharpal A (2021) Jack ma's ant group gets nod to operate consumer finance firm, a key step in fixing regulatory issues. *CNBC* .

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds., *Advances in Neural Information Processing Systems 25*, 1097–1105 (Curran Associates, Inc.), URL `http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf`.

Kumar A, Kumar N, Hussain M, Chaudhury S, Agarwal S (2014) Semantic clustering-based cross-domain recommendation. *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 137–141 (IEEE).

Lan Q, Xu X, Hu X (2017) Predictability of investment behavior based on personal characteristics about china's individual investors. *Eurasia Journal of Mathematics, Science and Technology Education* 13, URL `http://dx.doi.org/10.12973/ejmste/78104`.

Li P, Tuzhilin A (2020) Ddtcdr: Deep dual transfer cross domain recommendation. *Proceedings of the 13th International Conference on Web Search and Data Mining*, 331–339.

Liu B, Wei Y, Zhang Y, Yan Z, Yang Q (2018) Transferable contextual bandit for cross-domain recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Liu H, Hu Z, Mian A, Tian H, Zhu X (2014) A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-based systems* 56:156–166.

Liu J, Zhao P, Zhuang F, Liu Y, Sheng VS, Xu J, Zhou X, Xiong H (2020) Exploiting aesthetic preference in deep cross networks for cross-domain recommendation. *Proceedings of The Web Conference 2020*, 2768–2774.

Malmendier U (2021) Experience effects in finance: Foundations, applications, and future directions. Working Paper 29074, National Bureau of Economic Research, URL `http://dx.doi.org/10.3386/w29074`.

Man T, Shen H, Jin X, Cheng X (2017) Cross-domain recommendation: An embedding and mapping approach. *IJCAI*, volume 17, 2464–2470.

Margaris D, Vassilakis C, Spiliotopoulos D (2020) What makes a review a reliable rating in recommender systems? *Information Processing & Management* 57(6):102304.

Matz S, Kosinski M, Nave G, Stillwell D (2017) Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences* 114:201710966, URL `http://dx.doi.org/10.1073/pnas.1710966114`.

Mikolov T (2012) *Statistical language models based on neural networks*. Ph.D. thesis, Brno University of Technology.

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality (Curran Associates, Inc.), URL `http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf`.

Mnih A, Salakhutdinov RR (2008) Probabilistic matrix factorization. *Advances in neural information processing systems*, 1257–1264.

Nedelec T, Smirnova E, Vasile F (2017) Specializing joint representations for the task of product recommendation. *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*, 10–18, DLRS 2017 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450353533, URL `http://dx.doi.org/10.1145/3125486.3125489`.

Ng A, et al. (2011) Sparse autoencoder. *CS294A Lecture notes* 72(2011):1–19.

Nicholson N, Soane E, Fenton-O'Creevy M, Willman P (2005) Personality and domain-specific risk taking. *Journal of Risk Research* 8(2):157–176.

Oehler A, Wendt S, Wedlich F, Horn M (2018) Investors' personality influences investment decisions: Experimental evidence on extraversion and neuroticism. *Journal of Behavioral Finance* 19(1):30–48, URL `http://dx.doi.org/10.1080/15427560.2017.1366495`.

Oquab M, Bottou L, Laptev I, Sivic J (2014) Learning and transferring mid-level image representations using convolutional neural networks. *CVPR'14*, 1717–1724.

Oren B, Noam K (2016) Item2vec: Neural item embedding for collaborative filtering. *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)* 1–6.

Pan SJ, Yang Q (2010) A survey on transfer learning. *IEEE Trans. on Knowl. and Data Eng.* 22(10):1345–1359, ISSN 1041-4347, URL `http://dx.doi.org/10.1109/TKDE.2009.191`.

Pere S (2020) Covid-19 pandemic accelerated shift to e-commerce by 5 years, new report says. URL `https://techcrunch.com/2020/08/24/covid-19-pandemic-accelerated-shift-to-e-commerce-by-5-years-new-report-says/`.

Pompian MM, Longo JM (2004) A new paradigm for practical application of behavioral finance. *The Journal of Wealth Management* 7(2):9–15, ISSN 1534-7524, URL `http://dx.doi.org/10.3905/jwm.2004.434561`.

Rajendran DPD, Sundarraj RP (2021) Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings. *International Journal of Information Management Data Insights* 1(2):100027.

Sakar CO, Polat SO, Katircioglu M, Kastro Y (2019) Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and lstm recurrent neural networks. *Neural Computing and Applications* 31(10):6893–6908.

Schlosser AE, White TB, Lloyd SM (2006) Converting web site visitors into buyers: how web site investment increases consumer trusting beliefs and online purchase intentions. *Journal of marketing* 70(2):133–148.

ScrapeHero (2021) How many products does amazon sell? – march 2021. *ScrapeHero* URL `https://www.scrapehero.com/how-many-products-does-amazon-sell-march-2021/`.

Shapira B, Rokach L, Freilikhman S (2013) Facebook single and cross domain data for recommendation systems. *User Modeling and User-Adapted Interaction* 23(2-3):211–247.

Sharma P, Sivakumaran B, Marshall R (2010) Impulse buying and variety seeking: A trait-correlates perspective. *Journal of Business Research* 63(3):276–283.

Shi L, Zhao WX, Shen YD (2017) Local representative-based matrix factorization for cold-start recommendation. *ACM Transactions on Information Systems (TOIS)* 36(2):1–28.

Sicherman N, Loewenstein G, Seppi D, Utkus S (2016) Financial attention 29:863–897, URL `http://dx.doi.org/10.1093/rfs/hhv073`.

Singh AP, Gordon GJ (2008) Relational learning via collective matrix factorization. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 650–658.

Singh L, Singh S, Arora S, Borar S (2019) One embedding to do them all. *CoRR* abs/1906.12120, URL `http://arxiv.org/abs/1906.12120`.

Singh V, Hansen K, Blattberg R (2006) Market entry and consumer behavior: An investigation of a wal-mart supercenter. *Marketing Science* 25:457–476, URL `http://dx.doi.org/10.1287/mksc.1050.0176`.

Son LH (2016) Dealing with the new user cold-start problem in recommender systems: A comparative review. *Information Systems* 58:87–104.

Srebro N, Rennie JD, Jaakkola TS (2004) Maximum-margin matrix factorization. *NIPS*, volume 17, 1329–1336 (Citeseer).

Stachl C, Au Q, Schoedel R, Gosling SD, Harari GM, Buschek D, Völkel ST, Schuwerk T, Oldemeier M, Ullmann T, Hussmann H, Bischl B, Bühner M (2020) Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences* 117(30):17680–17687, ISSN 0027-8424, URL `http://dx.doi.org/10.1073/pnas.1920484117`.

Sun Q, Gu J, Yang B, Xu X, Xu R, Gao S, Liu H, Xu H (2021) Interest-oriented universal user representation via contrastive learning.

Tan S, Bu J, Qin X, Chen C, Cai D (2014) Cross domain recommendation based on multi-type media fusion. *Neurocomputing* 127:124–134.

Tan WK, Tang CY (2013) Does personality predict tourism information search and feedback behaviour? *Current issues in tourism* 16(4):388–406.

Taneja A, Arora A (2018) Cross domain recommendation using multidimensional tensor factorization. *Expert Systems with Applications* 92:304–316.

Tang J, Wu S, Sun J, Su H (2012) Cross-domain collaboration recommendation. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1285–1293.

Vyas S, Kumaranayake L (2006) Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning* 21(6):459–468, ISSN 0268-1080, URL `http://dx.doi.org/10.1093/heapol/czl029`.

Wang Z, Yang M (2019) How does investors' heterogeneous trust affect the complexity of financial products? a look into the development of online finance. *Managerial and Decision Economics* 40(4):425–438.

Wedel M, Zhang J (2004) Analyzing brand competition across subcategories. *Journal of Marketing Research* 41(4):448–456, URL `http://dx.doi.org/10.1509/jmkr.41.4.448.47017`.

Winkleby MA, Jatulis DE, Frank E, Fortmann SP (1992) Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *American Journal of Public Health* 82(6):816–820.

Winoto P, Tang T (2008) If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations. *New Generation Computing* 26(3):209–225.

Xia H, Hou Z (2016) Consumer use intention of online financial products: the yuebao example. *Financial Innovation* 2(1):1–12.

Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, 3320–3328, NIPS'14 (Cambridge, MA, USA: MIT Press).

Yu X, Jiang F, Du J, Gong D (2019) A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains. *Pattern Recognition* 94:96–109.

Yuan F, Yao L, Benatallah B (2019) Darec: Deep domain adaptation for cross-domain recommendation via transferring rating patterns. *arXiv preprint arXiv:1905.10760* .

Zhang Q, Hao P, Lu J, Zhang G (2019) Cross-domain recommendation with semantic correlation in tagging systems. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (IEEE).

Zhang Z, Jin X, Li L, Ding G, Yang Q (2016) Multi-domain active learning for recommendation. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2358–2364, AAAI'16 (AAAI Press).

Zhao C, Li C, Xiao R, Deng H, Sun A (2020) Catn: Cross-domain recommendation for cold-start users via aspect transfer network. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 229–238.

Zhou H, Liu J, Li Z, Yu J, Yang H (2021) Cross-domain user preference learning for cold-start recommendation. *arXiv preprint arXiv:2112.03667* .

Zhu F, Wang Y, Chen C, Liu G, Orgun M, Wu J (2020) A deep framework for cross-domain and cross-system recommendations. *arXiv preprint arXiv:2009.06215* .

Zhu F, Wang Y, Chen C, Zhou J, Li L, Liu G (2021) Cross-domain recommendation: challenges, progress, and prospects. *arXiv preprint arXiv:2103.01696* .

Zhu Y, Chen Y, Lu Z, Pan SJ, Xue GR, Yu Y, Yang Q (2011) Heterogeneous transfer learning for image classification. *Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Appendix

Table 8 Features Extracted from activities associated with product X, where X can represent Mutual Fund, Shampoo, Toothpaste, or Washer.

Extracted Features	Mutual Funds	Consumer Goods	Description	Type
X_NUNIQ	✔	✔	Number of unique item IDs clicked	Common Features
X_WEB	✔	✔	Total number of clicks from the Desktop site	
X_APP	✔	✔	Total number of clicks from the Mobile Site	
X_0-6	✔	✔	Number of clicks between the time range 0:00 to 6:00	
X_6-12	✔	✔	Number of clicks between the time range 6:00 to 12:00	
X_12-18	✔	✔	Number of clicks between the time range 12:00 to 18:00	
X_18-24	✔	✔	Number of clicks between the time range 18:00 to 24:00	
X_Risk_Level	✔		Number of order in risk level Z* (Z = 1, 2, 3, 4, 5)	Unique Mutual Funds Features
X_AReturn_less00	✔		Number of order with average annual return rate less than 0	
X_AReturn_less05	✔		Number of order with average annual return rate within [0, 0.5)	
X_AReturn_less10	✔		Number of order with average annual return rate within [0.5, 1)	
X_AReturn_less50	✔		Number of order with average annual return rate within [1, 5)	
X_CReturn_less00	✔		Number of order with average cumulative return rate less than 0	
X_CReturn_less50	✔		Number of order with average cumulative return rate within [0,5)	
X_CReturn_less100	✔		Number of order with average cumulative return rate within [5,10)	
X_CReturn_other	✔		Number of order with average cumulative return rate over 10	
X_MReturn_less00	✔		Number of order with average cumulative return rate less than 0	
X_MReturn_less10	✔		Number of order with average cumulative return rate within [0,1)	
X_MReturn_greater10	✔		Number of order with average cumulative return rate over 1	
X_PR_LEV_Z		✔	The number of orders in price level Z* (Z = 1, 2, 3, 4, 5)	Unique Consumer Goods Features
X_NUNIQ_BRAND		✔	The number of unique clicks by brand ID	
X_PR_ORDER		✔	The number of historical order	

(*Note: We discretize the prices by 20% percentile)

Table 9 Users Profile Features

User Profiles	Descriptions
Education	Education Degree (i.e., High School, Community College, Bachelor, Master or higher)
Profit Seeking	Prefer asset preservation / Substantial growth in assets
Risk Tolerance	No risk taking at all / Willing to take risk
Income	Income level (e.g., high income, stable income)

Table 10 Percentage of users corresponding with each feature in two clustered group

Feature		Low Socioeconomic Group	High Socioeconomic Group
Education	High School or Lower	91%	8%
	Community College	3%	20%
	Bachelor	2%	55%
	Master or PhD	2%	15%
Profit Seeking Appetite	Seek more profits	0.07%	77%
	Secure principles	5%	16%
	Asset Preservation	93%	0
	Substantial growth in assets	0.8%	6%
Risk Appetite	Willing to take some risks	0.07%	77%
	No care about low rate of return	5%	16%
	Do not want to take any risk	93%	0
	Willing to take huge risk	0.8%	6%
Income	Stable income, Family income and expenditure are basically balanced, Have short-term debt	4%	29%
	High income, Family income greater than expenditure, No large debt	2%	37%
	No stable income, borrowing from relatives	88%	0.3%
	Stable income, Family income and expenditure are balanced, Have long-term debt	3%	27%
	Stable income, Family income greater than expenditure, Have economic burden	1%	5%